

Mixing Properties of CSMA Networks on Partite Graphs

Alessandro Zocca
Dept. of Math. & Comp. Science
Eindhoven University of Technology
Den Dolech 2, 5612 AZ
Eindhoven, The Netherlands
a.zocca@tue.nl

Sem C. Borst
Dept. of Math. & Comp. Science
Eindhoven University of Technology
Den Dolech 2, 5612 AZ
Eindhoven, The Netherlands
s.c.borst@tue.nl

Johan S. H. van Leeuwen
Dept. of Math. & Comp. Science
Eindhoven University of Technology
Den Dolech 2, 5612 AZ
Eindhoven, The Netherlands
j.s.h.v.leeuwen@tue.nl

Abstract—We consider a stylized stochastic model for a wireless CSMA network. Experimental results in prior studies indicate that the model provides remarkably accurate throughput estimates for IEEE 802.11 systems. In particular, the model offers an explanation for the severe spatial unfairness in throughputs observed in such networks with asymmetric interference conditions. Even in symmetric scenarios, however, it may take a long time for the activity process to move between dominant states, giving rise to potential starvation issues.

In order to gain insight in the transient throughput characteristics and associated starvation effects, we examine in the present paper the behavior of the transition time between dominant activity states. We focus on partite interference graphs, and establish how the magnitude of the transition time scales with the activation rate and the sizes of the various network components. We also prove that in several cases the scaled transition time has an asymptotically exponential distribution as the activation rate grows large, and point out interesting connections with related exponentiality results for rare events and meta-stability phenomena in statistical physics. In addition, we investigate the convergence rate to equilibrium of the activity process in terms of mixing times.

Index Terms—CSMA networks, throughput starvation, interference graph, asymptotic exponentiality, mixing time, conductance.

I. INTRODUCTION AND MOTIVATION

We consider a stylized model for a network of nodes sharing a wireless medium according to a CSMA-type protocol. The network is described by an undirected graph (V, E) where the set of vertices V represents the various nodes of the network and the set of edges $E \subseteq V \times V$ indicates which pairs of nodes interfere. In other words, nodes that are neighbors in the interference graph are prevented from simultaneous activity, and thus the independent sets correspond to the feasible joint activity states. A node is said to be blocked whenever the node itself or any of its neighbors is active, and unblocked otherwise. Each node activates (starts a transmission) at an exponential rate ν whenever it is unblocked. The transmission durations of nodes are independent and exponentially distributed with unit mean.

This research was financially supported by The Netherlands Organization for Scientific Research (NWO) in the framework of the TOP-GO program and by an ERC Starting Grant.

Let $\Omega^* \subseteq \{0, 1\}^V$ be the collection of incidence vectors of the independent sets of the interference graph (V, E) , and let $X_t^* \in \Omega^*$ denote the joint activity state at time t , with its i -th element indicating whether or not node i is active at time t . Then $(X_t^*)_{t \geq 0}$ is a reversible Markov process [1] with stationary distribution

$$\pi_x(\nu) = \lim_{t \rightarrow \infty} \mathbb{P}\{X_t^* = x\} = \frac{\nu^{\|x\|_1}}{\sum_{y \in \Omega^*} \nu^{\|y\|_1}}, \quad x \in \Omega^*, \quad (1)$$

with $\|x\|_1 = \sum_{i \in V} x_i$ the number of active nodes in state x .

The above model has a long history [2]–[4]. It was rediscovered in the context of IEEE 802.11 systems in [5], and further developed in that setting in [6], [7]. While the modeling of the IEEE 802.11 back-off mechanism is less detailed than in the model of Bianchi [8], the general interference graph offers greater versatility and covers a broad range of topologies. Experimental results in [9] demonstrate that the model, while idealized, provides remarkably accurate throughput estimates for actual IEEE 802.11 systems. It is also worth observing that the model amounts to a special instance of a loss network [10], [11], and that the stationary distribution corresponds to the Gibbs measure of the hard-core model in statistical physics [12], [13].

An activity state in Ω^* is called *dominant* if it corresponds to an independent set of maximum size $\max_{x \in \Omega^*} \|x\|_1$. It follows from the stationary distribution (1) that only the dominant states retain probability mass as $\nu \rightarrow \infty$. This causes a severe problem, called *spatial unfairness*, when some nodes belong to fewer maximum-size independent sets than others as a result of asymmetric interference conditions [14], and receive far lower or even zero throughputs.

Even in symmetric scenarios where the long-term throughputs are equal, however, potential starvation issues can occur, since it may take a long time for the activity process to move between dominant states. Consider for example an interference graph where the nodes can be partitioned into two independent sets of maximum size. The activity process will spend roughly half of the time in each of the two associated dominant states as $\nu \rightarrow \infty$. Since each of the nodes is active in one of the dominant states, they will all receive equal long-term throughputs, so spatial unfairness is not an issue in

the long run. Yet, another source of severe unfairness arises as $\nu \rightarrow \infty$, because it will take an extremely long time for the activity process to move from one dominant state to the other (resembling *meta-stability phenomena* in statistical physics [15], [16]). As a result, each node will experience long sequences of transmissions in rapid succession, interspersed with extended periods of starvation.

In order to gain insight in the above issue, we explore in the present paper the behavior of the Markov process $(X_t^*)_{t \geq 0}$ in a scenario with a complete K -partite interference graph (V, E) . By this we mean that the nodes can be partitioned into K disjoint sets called *components*, such that two nodes block each other, if and only if they belong to different components. The key contributions of the present paper may be summarized as follows:

(i) We examine the asymptotic behavior (as $\nu \rightarrow \infty$) of the transition time between various activity states, exploiting classical results for absorption times in birth-and-death processes [17]–[19] and a representation of the transition time as a geometric random sum.

(ii) We establish that the magnitude of the transition time scales as $\Theta(\nu^{\beta-1})$, reflecting how ‘rigid’ the dominant activity pattern is as function of the activation rate ν and the exponent β , which is completely determined by the sizes of the various components.

(iii) We prove that in several cases the scaled transition time has an asymptotically exponential distribution and compare it with related exponentiality results for rare events [20]–[25].

(iv) We investigate the rate of convergence of the activity process to the equilibrium distribution, and demonstrate that the mixing time is of the same order as the escape time of the second-largest component.

The remainder of the paper is organized as follows. In Section II we present a model description and some preliminary results. In Section III we examine the asymptotic behavior of the transition time in the case of a bipartite graph. Section IV describes the extension of the results to arbitrary partite graphs. In Section V we investigate the rate of convergence to equilibrium in terms of mixing times. Section VI concludes with some remarks and a review of further extensions.

II. MODEL DESCRIPTION AND PRELIMINARY RESULTS

Consider the Markov process $(X_t^*)_{t \geq 0}$ as described in the introduction with a complete K -partite interference graph (V, E) . Thus the nodes in V can be partitioned into K disjoint sets called *components*, such that two nodes are connected by an edge in E , if and only if they belong to different components. In view of the symmetry, all the states with the same number of active nodes in a given component can be aggregated, and we only need to keep track of the number of active nodes, if any, and the component they belong to. This *state aggregation* yields an equivalent Markov process $(X_t)_{t \geq 0}$ on a star-shaped state space Ω with K branches which emanate from a common root node and correspond to the components of the interference graph. Specifically, $\Omega = \{0\} \cup \{(k, l) : 1 \leq l \leq L_k, 1 \leq k \leq K\}$, where the center

state 0 indicates that none of the nodes is active and state (k, l) corresponds to the situation where l nodes are active in the k -th component, denoted by C_k , with L_k denoting the size of C_k (see Figure 1). Notice that Ω can be alternatively described as $\{0\} \cup \bigcup_{k=1}^K C_k$. The transition rates of the process $(X_t)_{t \geq 0}$ are given by $q(0, (k, 1)) = L_k \nu$, $q((k, l), (k, l+1)) = (L_k - l) \nu$, $l = 1, \dots, L_k - 1$, $q((k, 1), 0) = 1$, and $q((k, l), (k, l-1)) = l$, $l = 2, \dots, L_k$, $k = 1, \dots, K$.

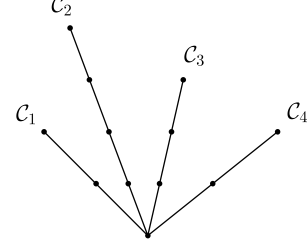


Fig. 1. Example of the state space Ω with $K = 4$.

The stationary distribution of the process $(X_t)_{t \geq 0}$ reads

$$\pi_0(\nu) = \left(1 + \sum_{k=1}^K \sum_{l=1}^{L_k} \binom{L_k}{l} \nu^l\right)^{-1}, \quad (2)$$

$$\pi_{(k,l)}(\nu) = \pi_0(\nu) \binom{L_k}{l} \nu^l, \quad l = 1, \dots, L_k, k = 1, \dots, K.$$

Denote by

$$T_{(k_1, l_1), (k_2, l_2)}(\nu) = \inf\{t > 0 : X_t = (k_2, l_2) | X_0 = (k_1, l_1)\}$$

the first-passage time of state (k_2, l_2) starting in (k_1, l_1) , which we also refer to as the *transition time* from state (k_1, l_1) to state (k_2, l_2) . In the next sections we will analyze the asymptotic behavior of the transition time as $\nu \rightarrow \infty$. As discussed in the introduction, the transition time provides useful insight in transient throughput characteristics and starvation effects in wireless CSMA networks.

In preparation for the asymptotic analysis of general scenarios, we first present a few results for the case where the two states (k_1, l_1) and (k_2, l_2) belong to the same component, i.e. $k_1 = k_2$ and $l_1 > l_2$. The presence of the other components is not relevant then for the transition time, and hence we focus on the case of just a single component ($K = 1$), and drop the component index for now. When $K = 1$, the process $(X_t)_{t \geq 0}$ evolves as an elementary birth-and-death process on the state space $\{L, L-1, \dots, 1, 0\}$, so that we can exploit several classical results for such processes.

A. Asymptotic growth rate

We first state how the expectation of the transition time $T_{l_1, l_2}(\nu)$ scales as $\nu \rightarrow \infty$. Here and throughout the paper we write $f(\nu) \sim g(\nu)$ to indicate that $\lim_{\nu \rightarrow \infty} f(\nu)/g(\nu) = 1$ as $\nu \rightarrow \infty$ for any two real-valued functions $f(\cdot)$ and $g(\cdot)$.

Proposition II.1. For $L \geq l_1 > l_2 \geq 0$,

$$\mathbb{E}T_{l_1, l_2}(\nu) \sim \frac{l_2!(L - l_2 - 1)!}{L!} \nu^{L - l_2 - 1}, \quad \nu \rightarrow \infty.$$

Proof: First observe that $\mathbb{E}T_{l_1, l_2}(\nu) = \sum_{l=l_1}^{l_2+1} \mathbb{E}T_{l, l-1}(\nu)$, so we can exploit a general result for birth-and-death pro-

cesses [18], which in the present case says that, for $0 < l \leq L$, $\mathbb{E}T_{l,l-1}(\nu) = \frac{1}{l} \sum_{n=l}^L \frac{\pi_n(\nu)}{\pi_l(\nu)}$. Now (2) implies that $\pi_n(\nu) = o(\pi_L(\nu))$ as $\nu \rightarrow \infty$ for all $n = l, \dots, L-1$, so that

$$\mathbb{E}T_{l,l-1}(\nu) \sim \frac{1}{l} \frac{\pi_L(\nu)}{\pi_l(\nu)} = \frac{(L-l)!(l-1)!}{L!} \nu^{L-l}, \quad \nu \rightarrow \infty.$$

Thus $\mathbb{E}T_{l,l-1}(\nu) = o(\mathbb{E}T_{l_2+1,l_2})$ as $\nu \rightarrow \infty$ for all $l = l_1, \dots, l_2$, and hence $\mathbb{E}T_{l_1,l_2}(\nu) \sim \mathbb{E}T_{l_2+1,l_2}(\nu)$ as $\nu \rightarrow \infty$. ■

In order to gain insight in starvation effects, we are particularly interested in the time for the activity process to reach the center state 0, referred to as *escape time*, because at such points in time nodes in other components have an opportunity to activate. Proposition II.1 shows that

$$\mathbb{E}T_{l,0}(\nu) \sim \frac{1}{L} \nu^{L-1}, \quad \nu \rightarrow \infty. \quad (3)$$

Hence, the escape time grows as a power of ν , where the exponent corresponds to the component size minus one, and is asymptotically not influenced by the starting state l_1 .

B. Asymptotic exponentiality

Having obtained the asymptotic growth rate, we now turn attention to the distribution of the scaled escape time, and show that it has an asymptotically exponential distribution. We will leverage the following well-known result for birth-and-death processes, which is commonly attributed to Keilson [19] or Karlin and McGregor [17].

Theorem II.2. *Consider a birth-and-death process with generator matrix Q on the state space $\{0, \dots, d\}$ started at state d . Assume that 0 is an absorbing state, and that the other birth rates $\{\lambda_i\}_{i=1}^{d-1}$ and death rates $\{\mu_i\}_{i=1}^d$ are positive. Then the absorption time in state 0 is distributed as the sum of d independent exponential random variables whose rate parameters are the d nonzero eigenvalues of $-Q$.*

Let $Q(\nu)$ be the generator matrix of the birth-and-death process $(X_t)_{t \geq 0}$ on the state space $\{L, L-1, \dots, 1, 0\}$, with 0 an absorbing state. Let $0 < \alpha_1(\nu) < \alpha_2(\nu) < \dots < \alpha_L(\nu)$ denote the non-zero eigenvalues of $-Q(\nu)$. The property that the eigenvalues $\{\alpha_i(\nu)\}_{i=1}^L$ are distinct, real and strictly positive, is well-known [26], see also the proof of Lemma II.3.

Applying Theorem II.2, we obtain that

$$T_{L,0}(\nu) \stackrel{d}{=} \sum_{i=1}^L Y_i(\nu), \quad (4)$$

where $Y_1(\nu), \dots, Y_L(\nu)$ are independent exponentially distributed random variables with $\mathbb{E}Y_i(\nu) = 1/\alpha_i(\nu)$.

Intuitively, the eigenvalue $\alpha_1(\nu)$ will become really small as $\nu \rightarrow \infty$, and so the mean escape time $\mathbb{E}T_{L,0}(\nu)$ will be dominated by $\mathbb{E}Y_1(\nu) = 1/\alpha_1(\nu)$. Combining (3) and (4) suggests that

$$\alpha_1(\nu) \sim \frac{1}{\mathbb{E}T_{L,0}(\nu)} \sim \frac{L}{\nu^{L-1}}, \quad \nu \rightarrow \infty.$$

Indeed, we can make this precise with the following result for the growth rates of the eigenvalues as $\nu \rightarrow \infty$ and their relation to the mean escape time $\mathbb{E}T_{L,0}(\nu)$. The proof of

the result is presented in Appendix A, and exploits detailed information about the growth rates of the eigenvalues obtained via symmetrization and the Gershgorin circle theorem.

Lemma II.3.

$$\lim_{\nu \rightarrow \infty} \alpha_i(\nu) \cdot \mathbb{E}T_{L,0}(\nu) = \begin{cases} 1, & i = 1, \\ \infty, & i = 2, \dots, L. \end{cases}$$

The above lemma shows that the smallest eigenvalue $\alpha_1(\nu)$ becomes dominant as $\nu \rightarrow \infty$, but also proves the asymptotic exponentiality of the escape time. Indeed (4) means that the Laplace transform of the scaled escape time is

$$\mathcal{L}_{T_{L,0}(\nu)/\mathbb{E}T_{L,0}(\nu)}(s) = \prod_{i=1}^L \left(1 + \frac{s}{\alpha_i(\nu) \cdot \mathbb{E}T_{L,0}(\nu)}\right)^{-1}.$$

Lemma II.3 then implies that

$$\lim_{\nu \rightarrow \infty} \mathcal{L}_{T_{L,0}(\nu)/\mathbb{E}T_{L,0}(\nu)}(s) = \frac{1}{1+s}.$$

The continuity theorem for Laplace transforms then yields that the scaled escape time has an asymptotically exponential distribution as stated in the next theorem, where $\text{Exp}(\lambda)$ denotes an exponentially distributed random variable with mean $1/\lambda$.

Theorem II.4. $\frac{T_{L,0}(\nu)}{\mathbb{E}T_{L,0}(\nu)} \xrightarrow{d} \text{Exp}(1), \quad \nu \rightarrow \infty.$

The above result may be interpreted as follows. For large ν , state L is frequently visited, while state 0 becomes rare. This suggests that the probability of hitting state 0 before the first return to state L becomes small. So the time $T_{L,0}(\nu)$ consists of a geometric number of excursions from L which return to L without hitting 0, followed by part of the excursion that hits 0. Hence, apart from this final excursion, $T_{L,0}(\nu)$ is the sum of a large geometric number of i.i.d. random variables, which should be asymptotically exponential.

The fact that the time until the first occurrence of a rare event is asymptotically exponential, is a widely observed phenomenon in probability [25]. In order to establish exponentiality of hitting some subset B of the state space, one typically decomposes the process into regenerative cycles, and assumes that (i) the probability of hitting B in a single cycle is small, and (ii) the length of the cycle in which B is hit is asymptotically negligible compared with the mean cycle length [23], [25]. For the case $K = 1$, both assumptions hold, and an alternative proof of Theorem II.4 can be obtained using [23, Thm.1] (which is a generalized version of the theorem proved in [24]). However, this general theorem for regenerative processes will fail for the majority of cases considered in this paper. For instance, in the case $K = 2$ and $L_1 = L_2$ (two identical components), the process $(X_t)_{t \geq 0}$ will exhibit *bistable behavior*, in the sense that, as $\nu \rightarrow \infty$, it spends extremely long periods circling around either one of the leaf nodes. For analyzing the transition time between the two dominant leaf nodes, assumption (i) remains valid, because it is still extremely hard to go from one leaf to the other, but assumption (ii) will be violated. Indeed, when the process crosses all the way from one leaf to the other, the

length of the latter cycle will be of the same order as the entire hitting time. Despite the fact that the regenerative approach no longer works, we will prove (using different methods) in the next section that the scaled transition time between dominant states is still asymptotically exponential.

Let us finally remark that for reversible Markov chains similar exponentiality results were established in [20]–[22]. Aldous [20] showed that a result like Theorem II.4 can be expected when the underlying Markov process converges rapidly to stationarity. Indeed, the Markov process $(X_t)_{t \geq 0}$ for the case $K = 1$ turns out to have a small mixing time. However, the picture changes drastically in the case $K \geq 2$, for which we prove in Section V that the mixing time is $\Theta(\nu^{\beta-1})$ with β the size of the second-largest component; hence the mixing time is of the same order as the escape time of the second-longest branch, see (3). For large ν , this implies that the Markov process is extremely slowly mixing, which is another way of understanding why the regenerative approach sketched above fails in this situation.

III. BIPARTITE INTERFERENCE GRAPH

In this section we examine the asymptotic behavior of the transition time $T_{(k_1, l_1), (k_2, l_2)}(\nu)$ as $\nu \rightarrow \infty$ for any pair of states (k_1, l_1) and (k_2, l_2) when the interference graph is bipartite. We will establish how the expectation scales (Theorem III.1) and use two different methods to prove that the scaled transition time has an asymptotically exponential distribution when states (k_1, l_1) and (k_2, l_2) belong to different branches (Theorem III.2). Most of these results in fact extend to arbitrary partite graphs, as will be shown in the next section. However, we treat the case of a bipartite graph separately first, since it allows us to develop the key ideas in a relatively transparent setting, while being sufficiently rich to exhibit the essential qualitative characteristics of general scenarios.

Consider a bipartite interference graph, i.e. $K = 2$, where the two components \mathcal{C}_1 and \mathcal{C}_2 have sizes L_1 and L_2 , respectively. In this case, the two branches of the state space along with the root node form a simple linear array. For notational convenience, we will therefore relabel the states, indicating the number of active nodes in \mathcal{C}_1 by a negative integer in $\{-L_1, \dots, -1\}$ and representing the number of those in \mathcal{C}_2 by a positive integer in $\{1, \dots, L_2\}$.

The activity process $(X_t)_{t \geq 0}$ on $\Omega = \mathcal{C}_1 \cup \{0\} \cup \mathcal{C}_2$ evolves as a birth-and-death process with transition rates

$$q(l, l+1) = \begin{cases} |l| & \text{if } -L_1 \leq l < 0, \\ (L_2 - l)\nu & \text{if } 0 \leq l < L_2, \end{cases}$$

$$q(l, l-1) = \begin{cases} (L_1 - |l|)\nu & \text{if } -L_1 < l \leq 0, \\ l & \text{if } 0 < l \leq L_2, \end{cases}$$

yielding

$$\pi_{-l}(\nu) = \pi_0(\nu) \binom{L_1}{l} \nu^l, \quad l = 1, \dots, L_1, \quad (5)$$

$$\pi_m(\nu) = \pi_0(\nu) \binom{L_2}{m} \nu^m, \quad m = 1, \dots, L_2. \quad (6)$$

A. Asymptotic growth rate

We first determine how the expectation of the transition time $T_{-l_1, l_2}(\nu)$ scales as $\nu \rightarrow \infty$.

Theorem III.1. *For any $0 < l_1 \leq L_1$ and $0 < l_2 \leq L_2$,*

$$\mathbb{E}T_{-l_1, l_2}(\nu) \sim \frac{L_1 + L_2}{L_1 L_2} \nu^{L_1-1}, \quad \nu \rightarrow \infty. \quad (7)$$

In the CSMA context, the above theorem implies that when at least one of the nodes in component k is currently active, $k = 1, 2$, it will approximately take an amount of time $\frac{L_1 + L_2}{L_1 L_2} \nu^{L_k-1}$ before any number of nodes in the other component $3 - k$ will have a chance to transmit.

Proof: As before, note that for any $l_1 < l_2$ $\mathbb{E}T_{l_1, l_2}(\nu) = \sum_{l=l_1}^{l_2-1} \mathbb{E}T_{l, l+1}(\nu)$, so we can again use a general result for birth-and-death processes [18], which in the present case says that $\mathbb{E}T_{l, l+1}(\nu) = \frac{1}{q(l, l+1)} \sum_{n=-L_1}^l \frac{\pi_n(\nu)}{\pi_l(\nu)}$, for $-L_1 \leq l < L_2$. Now (5) and (6) imply that $\pi_n(\nu) = o(\pi_{-L_1}(\nu))$ as $\nu \rightarrow \infty$ for all $n = -L_1 + 1, \dots, \min\{L_1 - 1, L_2\}$ and $\pi_n(\nu) = o(\pi_l(\nu))$ as $\nu \rightarrow \infty$ for all $n = -L_1, \dots, l - 1$, when $l = L_1 + 1, \dots, L_2$. Thus, for $-L_1 \leq l < 0$, as $\nu \rightarrow \infty$,

$$\mathbb{E}T_{l, l+1}(\nu) \sim \frac{1}{l} \frac{\pi_{-L_1}(\nu)}{\pi_l(\nu)} = \frac{(L_1 - |l|)!(|l| - 1)!}{L_1!} \nu^{L_1 - |l|},$$

and, for $0 \leq l \leq L_2$, as $\nu \rightarrow \infty$, $\mathbb{E}T_{l, l+1}(\nu)$ scales as

$$\begin{cases} \frac{\nu^{-1}}{(L_2 - l)} \frac{\pi_{-L_1}(\nu)}{\pi_l(\nu)} = \frac{(L_2 - l - 1)!}{L_2!} \nu^{L_1 - l - 1} & \text{if } l < L_1, \\ \frac{\nu^{-1}}{(L_2 - L_1)} \frac{\pi_{-L_1}(\nu) + \pi_{L_1}(\nu)}{\pi_{L_1}(\nu)} = \frac{L_2! + L_1!(L_2 - L_1)!}{(L_2 - L_1)L_2!} \nu^{-1} & \text{if } l = L_1, \\ \frac{\nu^{-1}}{(L_2 - l)} \frac{\pi_l(\nu)}{\pi_l(\nu)} = \frac{1}{L_2 - l} \nu^{-1} & \text{if } l > L_1. \end{cases} \quad (8)$$

It follows that $\mathbb{E}T_{l, l+1}(\nu) = o(\mathbb{E}T_{-1, 0}(\nu))$ as $\nu \rightarrow \infty$ for all $l = -L_1, \dots, -2$ and $l = 1, \dots, L_2 - 1$. Hence, as $\nu \rightarrow \infty$,

$$\mathbb{E}T_{-l_1, l_2}(\nu) \sim \mathbb{E}T_{-1, 0}(\nu) + \mathbb{E}T_{0, 1}(\nu) \sim \frac{1}{L_1} \nu^{L_1-1} + \frac{1}{L_2} \nu^{L_1-1},$$

which yields (7). \blacksquare

Remark 1. Theorem III.1 only deals with the case where the two states $-l_1$ and l_2 belong to different components, but extends to the case where the two states belong to the same component. In case $0 \leq l_2 < l_1 \leq L_1$, the transition time $T_{-l_1, -l_2}(\nu)$ is not influenced by the presence of the second component, and its expected value thus follows from Proposition II.1. In case $0 \leq l_1 < l_2 \leq L_2$, it may be deduced from (8) that $\mathbb{E}T_{l_1, l_2}(\nu) \sim \mathbb{E}T_{l_1, l_1+1}(\nu)$ as $\nu \rightarrow \infty$, yielding

$$\mathbb{E}T_{l_1, l_2}(\nu) \sim \begin{cases} \frac{(L_2 - l_1 - 1)!}{L_2!} \nu^{L_1 - l_1 - 1} & \text{if } l_1 < L_1 \\ \frac{L_2! + L_1!(L_2 - L_1)!}{(L_2 - L_1)L_2!} \nu^{-1} & \text{if } l_1 = L_1, \nu \rightarrow \infty. \\ \frac{1}{L_2 - l_1} \nu^{-1} & \text{if } l_1 > L_1 \end{cases}$$

B. Asymptotic exponentiality

In the previous subsection we obtained the asymptotic growth rate of the mean transition time $\mathbb{E}T_{-l_1, l_2}(\nu)$ as $\nu \rightarrow \infty$. We now turn attention to the scaled transition time and will prove that it has an asymptotically exponential distribution as stated in the next theorem.

Theorem III.2. For any $0 < l_1 \leq L_1$ and $0 < l_2 \leq L_2$

$$\frac{T_{-l_1, l_2}(\nu)}{\mathbb{E}T_{-l_1, l_2}(\nu)} \xrightarrow{d} \text{Exp}(1), \quad \nu \rightarrow \infty.$$

Before providing a proof, we first present an interpretation of the above theorem. In order for the process to make a transition from state $-l_1$ to state l_2 , it must first reach state 0. Proposition II.1 indicates that the expected transition time from state $-L_1$ to state $-l_1$ is asymptotically negligible compared to the expected transition time from state $-l_1$ to state 0. This means that the transition time from state $-l_1$ to state 0 asymptotically behaves as the transition time from state $-L_1$ to state 0. Theorem II.4 shows that the latter time, after scaling, has an asymptotically exponential distribution. Once the process has reached state 0 for the first time, there are two possible scenarios. With probability $p = L_1/(L_1 + L_2)$, the process returns to state -1 , and then almost surely falls back to state $-L_1$ quite rapidly (compared with the time to reach state 0). With probability $1 - p$, the process moves to state 1 , and then most likely is attracted to state l_2 quite quickly.

In conclusion, in order for the process to make a transition from state $-l_1$ to state l_2 it must visit state 0 once plus a geometrically distributed number of times with parameter p . The successive time periods to reach state 0 are independent and, after scaling, exponentially distributed. Observing that the sum of one plus a geometrically distributed number of independent and exponentially distributed random variables is again exponentially distributed, we arrive at the statement of the above theorem.

Obviously, the above argumentation is heuristic, and in the next two subsections we will provide a rigorous proof. The first proof method is analytical in nature, and relies on an asymptotic characterization of the eigenvalues of the generator matrix, while the second proof is more probabilistic, and in fact closely mirrors the above intuitive explanation.

Remark 2. Theorem III.2 only deals with the case where the two states $-l_1$ and l_2 belong to different components, but extends to the case where the two states belong to the same component. In case $0 \leq l_2 < l_1 \leq L_1$, the transition time $T_{-l_1, -l_2}(\nu)$ is not affected by the presence of the second component, and thus has an asymptotically exponential distribution according to Theorem II.4. In contrast, in case $0 \leq l_1 < l_2 \leq L_2$, a fundamentally different situation emerges. In that case, a series of upward transitions from state l_1 to state l_2 occurs in rapid succession with high probability, so that the transition time converges in distribution to 0, even though its expectation may not tend to 0 and in fact grow to infinity as $\nu \rightarrow \infty$ when $l_1 < L_1$, as observed in Remark 1.

C. Analytical approach

We now extend the analytical approach of Section II to prove the asymptotic exponentiality of the scaled transition time as stated in Theorem III.2. For convenience, we restrict attention to $l_1 = L_1$, but the proof readily extends to any $l_1 \in \{1, \dots, L_1 - 1\}$.

Let $Q(\nu)$ be the generator matrix of the birth-and-death process $(X_t)_{t \geq 0}$ on the state space $\{-L_1, \dots, -1, 0, 1, \dots, l_2\}$ with l_2 an absorbing state. Let $0 < \gamma_1(\nu) < \gamma_2(\nu) < \dots < \gamma_{L_1+l_2}(\nu)$ denote the non-zero eigenvalues of $-Q(\nu)$.

Theorem II.2 implies that $T_{-L_1, l_2}(\nu) \stackrel{d}{=} \sum_{i=1}^{L_1+l_2} Y_i(\nu)$, where $Y_1(\nu), \dots, Y_{L_1+l_2}(\nu)$ are independent and exponentially distributed random variables with $\mathbb{E}Y_i(\nu) = 1/\gamma_i(\nu)$.

The following lemma shows that the smallest eigenvalue $\gamma_1(\nu)$ becomes dominant as $\nu \rightarrow \infty$, and implies the asymptotic exponentiality stated in Theorem III.2. The proof is similar to that of Lemma II.3 and thus omitted.

Lemma III.3.

$$\lim_{\nu \rightarrow \infty} \gamma_i(\nu) \cdot \mathbb{E}T_{-L_1, l_2}(\nu) = \begin{cases} 1, & i = 1, \\ \infty, & i = 2, \dots, L_1 + l_2. \end{cases}$$

D. Probabilistic approach

We now present an alternative, probabilistic approach to establish the asymptotic exponentiality of the scaled transition time as stated in Theorem III.2. In contrast to the analytical method as used in the previous subsection, the probabilistic approach in fact extends to arbitrary partite graphs, but we first focus on the case of a bipartite graph in order to illuminate the key ideas. The approach relies on a stochastic decomposition of the transition time into independent random variables which are easier to handle. In order to obtain the stochastic decomposition, we consider the evolution of the process as it makes a transition from state $-l_1$ to a state l_2 , and define the following random variables:

- $T_{-l_1, -1}(\nu)$: time to reach state -1 for the first time;
- $T_{-1, 0}^{(0)}(\nu)$: time to reach state 0 once the process has reached state -1 for the first time;
- N : number of times the process makes a transition $0 \rightarrow -1$ before the first transition $0 \rightarrow 1$ occurs;
- $\hat{T}_{0, -1}^{(i)}(\nu)$: time spent in state 0 before the i -th transition back to state -1 , $i = 1, \dots, N$;
- $\hat{T}_{0, 1}(\nu)$: time spent in state 0 before the first transition to state 1;
- $T_{-1, 0}^{(i)}(\nu)$: time to return to state 0 after the i -th transition back to state -1 , $i = 1, \dots, N$;
- $T_{1, l_2}(\nu)$: time to reach state l_2 once the process has reached state 1 for the first time.

By definition, the transition time may be represented as

$$T_{-L_1, l_2} \stackrel{d}{=} T_{-L_1, -1} + T_{-1, 0}^{(0)} + \sum_{i=1}^N \left(\hat{T}_{0, -1}^{(i)} + T_{-1, 0}^{(i)} \right) + \hat{T}_{0, 1} + T_{1, l_2}, \quad (9)$$

where the dependence on the parameter ν is suppressed for compactness. We can make the following observations:

- The random variables $T_{-L_1, -1}$ and T_{1, l_2} are distributed as typical hitting times for the respective pairs of states;
- The random variables $T_{-1, 0}^{(i)}$, $i = 0, 1, \dots, N$, are i.i.d. copies of a typical transition time $T_{-1, 0}$;
- The random variables $\hat{T}_{0, 1}$ and $\hat{T}_{0, -1}^{(i)}$, $i = 1, 2, \dots, N$, are i.i.d. copies of a random variable $T_0 \stackrel{d}{=} \text{Exp}((L_1 + L_2)\nu)$,

which represents the residence time in state 0;

- $N \stackrel{d}{=} \text{Geo}(p)$, $p = \frac{L_1}{L_1 + L_2}$, independent of the parameter ν ;
- All the random variables are independent.

Based on the stochastic representation in (9) and the above observations, we now proceed to give a proof of the asymptotic exponentiality of the transition time as stated in Theorem III.2. The proof consists of three main parts: (i) the first part establishes that the distribution of the scaled transition time $T_{-l_1, l_2}(\nu)$ asymptotically coincides with that of a dominant term $U(\nu)$, which involves a random sum of i.i.d. random variables (Proposition III.4); (ii) the second part serves to identify the asymptotic behavior of each of these random variables (Proposition III.5); (iii) the third part then shows that a scaled random sum asymptotically behaves as the random sum of the scaled terms (Proposition III.6).

Part (i): According to Theorem III.1 and Remark 1, the mean values of the random variables $T_{-1,0}^{(i)}$, $i = 0, 1, \dots, N$, asymptotically dominate, i.e. they are an order-of-magnitude larger than those of all other random variables in (9) as $\nu \rightarrow \infty$. This suggests that the asymptotic behavior of the transition time $T_{-l_1, l_2}(\nu)$ will be determined by that of $U(\nu) = \sum_{i=0}^N T_{-1,0}^{(i)}(\nu)$, and in particular that the distribution of the scaled transition time $T_{-l_1, l_2}(\nu)/\mathbb{E}T_{-l_1, l_2}(\nu)$ asymptotically coincides, as $\nu \rightarrow \infty$, with that of $U(\nu)/\mathbb{E}U(\nu)$. This follows as a special case of Proposition III.4 below, whose proof is technical, but relatively straightforward, and omitted because of page limitations. In passing we note that the above observations also imply that the expectation of $T_{-l_1, l_2}(\nu)$ scales as $\mathbb{E}U(\nu) = (1 + \mathbb{E}N(\nu)) \mathbb{E}T_{-1,0}(\nu) = \frac{L_1 + L_2}{L_2} \mathbb{E}T_{-1,0}(\nu)$, with $\mathbb{E}T_{-1,0}(\nu) \sim \nu^{L_1-1}/L_1$ as $\nu \rightarrow \infty$ according to Proposition II.1, which corroborates Theorem III.1.

Part (ii): In order to determine the asymptotic behavior of the random variable $T_{-1,0}(\nu)$, we first observe that $T_{-L_1,0}(\nu) \stackrel{d}{=} T_{-L_1,-1}(\nu) + T_{-1,0}(\nu)$. According to Theorem III.1, the mean value of $T_{-L_1,-1}(\nu)$ is asymptotically negligible compared to that of $T_{-1,0}(\nu)$, which suggests that the asymptotic behavior of $T_{-1,0}(\nu)$ is equivalent to that of $T_{-L_1,0}(\nu)$. Since Theorem II.4 states that $T_{-L_1,0}(\nu)/\mathbb{E}T_{-L_1,0}(\nu)$ has an asymptotically exponential distribution, this would imply that the same holds for the scaled hitting time $T_{-1,0}(\nu)/\mathbb{E}T_{-1,0}(\nu)$, as is covered as a special case of Proposition III.5, stated and proved below.

Part (iii): It remains to be shown that $U(\nu)/\mathbb{E}U(\nu)$, with $U(\nu) = \sum_{i=0}^N T_{-1,0}^{(i)}(\nu)$, asymptotically behaves as the random variable $\frac{1}{1+\mathbb{E}N} \sum_{i=0}^N Y_i$, when $T_{-1,0}(\nu)/\mathbb{E}T_{-1,0}(\nu) \xrightarrow{d} Y$, and Y_1, Y_2, \dots are i.i.d. copies of the random variable Y . This follows as a special case of Proposition III.6 below, whose proof is in Appendix B. In particular, if N is geometrically distributed and Y is exponentially distributed, then $U(\nu)/\mathbb{E}U(\nu)$ has an asymptotically exponential distribution as well.

Proposition III.4. *Let $T(\nu), U(\nu), V(\nu), W(\nu)$ be non-negative random variables such that*

- (i) $\lim_{\nu \rightarrow \infty} \mathbb{E}V(\nu)/\mathbb{E}U(\nu) = 0 = \lim_{\nu \rightarrow \infty} \mathbb{E}W(\nu)/\mathbb{E}U(\nu)$;
- (ii) *For every $\nu > 0$, $U - V \leq_{\text{st}} T \leq_{\text{st}} U + W$, i.e. $\forall t > 0$*

$$\mathbb{P}\{U - V > t\} \leq \mathbb{P}\{T > t\} \leq \mathbb{P}\{U + W > t\};$$

(iii) $U(\nu)/\mathbb{E}U(\nu) \xrightarrow{d} X$ as $\nu \rightarrow \infty$, where X is a random variable independent of ν with continuous c.d.f..

Then $T(\nu)/\mathbb{E}T(\nu) \xrightarrow{d} X$ as $\nu \rightarrow \infty$.

Proposition III.5. *For any $0 < l \leq L$,*

$$\frac{T_{l,0}(\nu)}{\mathbb{E}T_{l,0}(\nu)} \xrightarrow{d} \text{Exp}(1), \quad \nu \rightarrow \infty.$$

Proof: The birth-and-death structure of the process and the strong Markov property yield the stochastic identity $T_{L,0}(\nu) \stackrel{d}{=} T_{L,l}(\nu) + T_{l,0}(\nu)$, which gives the stochastic bounds $T_{L,0}(\nu) - T_{L,l}(\nu) \leq_{\text{st}} T_{l,0}(\nu) \leq_{\text{st}} T_{L,0}(\nu)$ (the two terms in the lower bound being dependent). It follows from Theorem II.4 that $T_{L,0}(\nu)/\mathbb{E}T_{L,0}(\nu) \xrightarrow{d} \text{Exp}(1)$ as $\nu \rightarrow \infty$. In order to complete the proof, we can then use Proposition III.4, taking $U(\nu) = T_{L,0}(\nu)$, $V(\nu) = T_{L,l}(\nu)$ and $W(\nu) = 0$. The condition which needs to be checked is $\lim_{\nu \rightarrow \infty} \frac{\mathbb{E}V(\nu)}{\mathbb{E}U(\nu)} = 0$, which follows directly from Proposition II.1. ■

Proposition III.6. *For every $\nu > 0$, define $S_M(\nu) := \sum_{i=1}^M X_i(\nu)$, where M is an integer-valued random variable and $\{X_i(\nu)\}_{i \geq 1}$ is a sequence, independent of M , of i.i.d. copies of a random variable $X(\nu)$, with $\mathbb{E}X(\nu) < \infty$. Assume that $X(\nu)/\mathbb{E}X(\nu) \xrightarrow{d} Y$ as $\nu \rightarrow \infty$, where Y is some unit-mean random variable. Then*

$$\frac{S_M(\nu)}{\mathbb{E}S_M(\nu)} \xrightarrow{d} \frac{1}{\mathbb{E}M} \sum_{i=1}^M Y_i, \quad \nu \rightarrow \infty,$$

where $\{Y_i\}_{i \geq 1}$ is a sequence, independent of M , of i.i.d. copies of the random variable Y .

IV. ARBITRARY PARTITE GRAPHS

In this section we investigate the asymptotic behavior of the transition time $T_{(k_1, l_1), (k_2, l_2)}(\nu)$ as $\nu \rightarrow \infty$ for any pair of states (k_1, l_1) and (k_2, l_2) in arbitrary partite graphs. While the key ideas are similar to those used for the bipartite graph in the previous section, there are now some dependencies that require specific treatment and some different qualitative features that arise in certain scenarios. In particular, it turns out that the scaled transition time may no longer have an asymptotically exponential distribution. This scenario arises when there are longer branches than k_1 , in which case the number of returns to the root node that dominate the transition time is no longer geometrically distributed plus one, but just geometrically distributed.

As in the case of the bipartite graph, the proof approach involves a stochastic representation of the transition time, but some of the terms are now no longer entirely independent. In order to obtain the stochastic representation, we consider the evolution of the process as it makes a transition from a state (k_1, l_1) to a state (k_2, l_2) , and define the following random variables:

- $T_{(k_1, l_1), (k_1, 1)}(\nu)$: time to reach state $(k_1, 1)$ for the first time;
- $T_{(k_1, 1), 0}^{(0)}(\nu)$: time to reach state 0 after state $(k_1, 1)$ is visited for the first time;

- N_k : number of times the process makes a transition $0 \rightarrow (k, 1)$, $k \neq k_2$, before the first transition $0 \rightarrow (k_2, 1)$ occurs;
- $\hat{T}_{0,(k,1)}^{(i)}(\nu)$: time spent in state 0 before the i -th transition back to state $(k, 1)$, $k \neq k_2$, $i = 1, \dots, N_k$;
- $\hat{T}_{0,(k_2,1)}(\nu)$: time spent in state 0 before the first transition to state $(k_2, 1)$;
- $T_{(k,1),0}^{(i)}(\nu)$: time to return to state 0 after the i -th transition back to state $(k, 1)$, $k \neq k_2$, $i = 1, \dots, N_k$;
- $T_{(k_2,1),(k_2,l_2)}(\nu)$: time to reach state (k_2, l_2) once the process has reached state $(k_2, 1)$ for the first time.

By definition, the transition time may be represented as

$$T_{(k_1,l_1),(k_2,l_2)} \stackrel{d}{=} T_{(k_1,l_1),(k_1,1)} + T_{(k_1,1),0}^{(0)} \quad (10)$$

$$+ \sum_{k \neq k_2} \sum_{i=1}^{N_k} \left(\hat{T}_{0,(k,1)}^{(i)} + T_{(k,1),0}^{(i)} \right) + \hat{T}_{0,(k_2,1)} + T_{(k_2,1),(k_2,l_2)},$$

where the dependence on the parameter ν is suppressed for compactness. Define $L := \sum_{k=1}^K L_k$, $\gamma := L\nu$, and $p_k := L_k/L$ for all $k = 1, \dots, K$.

We can make the following observations:

- The random variables $T_{(k_1,l_1),(k_1,1)}$ and $T_{(k_2,1),(k_2,l_2)}$ are distributed as typical transition times for the respective pairs of states;
- The random variables $T_{(k,1),0}^{(i)}$ are i.i.d. copies of a typical transition time $T_{(k,1),0}$, $i = 0, \dots, N_k$, $k \neq k_2$;
- The random variables $\hat{T}_{0,(k_2,1)}$ and $\hat{T}_{0,(k,1)}^{(i)}$, $k \neq k_2$, $i = 1, \dots, N_k$, are i.i.d. copies of a random variable $T_0 \stackrel{d}{=} \text{Exp}(\gamma)$, which is the residence time in state 0;
- $N := \sum_{k \neq k_2} N_k \stackrel{d}{=} \text{Geo}(1 - p_{k_2})$, independent of the parameter ν ;
- Given $N = n$, $\bar{N} = (N_1, \dots, N_{k_2-1}, N_{k_2+1}, \dots, N_K)$ has a multinomial distribution with parameters n and $\bar{p}_1, \dots, \bar{p}_{k_2-1}, \bar{p}_{k_2+1}, \dots, \bar{p}_K$, with $\bar{p}_k = p_k/(1 - p_{k_2})$

$$\begin{aligned} & \mathbb{P}\{\bar{N} = (n_1, \dots, n_{k_2-1}, n_{k_2+1}, \dots, n_K)\} \\ &= p_{k_2}(1 - p_{k_2})^{\sum_{k \neq k_2} n_k} \binom{\sum_{k \neq k_2} n_k}{n_1, \dots, n_K} \prod_{k \neq k_2} \bar{p}_k^{n_k} \\ &= p_{k_2} \binom{\sum_{k \neq k_2} n_k}{n_1, \dots, n_K} \prod_{k \neq k_2} p_k^{n_k}; \end{aligned} \quad (11)$$

- All the random variables representing time durations are mutually independent, as well as independent of the random variables N_k , $k \neq k_2$.

Define $L_* := \max_{j \neq k_2} L_j$, $K_* := \{k \neq k_2 : L_k = L_*\}$, and $p_* := |K_*|L_*/(|K_*|L_* + L_{k_2})$.

We first use the stochastic representation (10) to establish how the expectation of the transition time scales.

Theorem IV.1. For $0 < l_1 \leq L_{k_1}$, $0 < l_2 \leq L_{k_2}$, $k_1 \neq k_2$,

$$\mathbb{E}T_{(k_1,l_1),(k_2,l_2)}(\nu) \sim \left(\frac{I_{\{k_1 \in K_*\}}}{L_*} + \frac{|K_*|}{L_{k_2}} \right) \nu^{L_*-1}, \quad \nu \rightarrow \infty.$$

In the CSMA context, the above theorem implies that for any component k when some other component is presently active, it will roughly take an amount of time of the order ν^{L_*} , $L_* = \max_{j \neq k} L_j$ before any number of nodes in component k

will get an opportunity to transmit. The proof of the above theorem is subsumed in that of the next one.

We now proceed to determine the asymptotic distribution of the scaled transition time.

Theorem IV.2. For $0 < l_1 \leq L_{k_1}$, $0 < l_2 \leq L_{k_2}$, $k_1 \neq k_2$,

$$\frac{T_{(k_1,l_1),(k_2,l_2)}(\nu)}{\mathbb{E}T_{(k_1,l_1),(k_2,l_2)}(\nu)} \xrightarrow{d} \frac{1}{\mathbb{E}M} \sum_{i=1}^M Y_i, \quad \nu \rightarrow \infty,$$

where $M \stackrel{d}{=} \text{Geo}(p_*) + I_{\{k_1 \in K_*\}}$ and Y_i are independent exponentially distributed random variables with unit mean.

Proof: The proof is similar to that for the bipartite graph and is based on the stochastic representation (10) of the transition time $T_{(k_1,l_1),(k_2,l_2)}(\nu)$. Proposition II.1 implies that $\mathbb{E}T_{(k_1,l_1),(k_1,1)}(\nu) \sim \nu^{L_{k_1}-2}$ and $\mathbb{E}T_{(k_1,1),0}(\nu) \sim \nu^{L_k-1}$ for every branch k . Moreover, $\mathbb{E}\hat{T}_{0,(k,1)}(\nu) = o(1)$, and it is easily verified that $\mathbb{E}T_{(k_2,1),(k_2,l_2)}(\nu) = o(\nu^{L_*-1})$ similarly as in the proof of Theorem III.1. Thus the asymptotically dominant term in (10) is

$$T_{(k_1,1),0}(\nu)I_{\{k_1 \in K_*\}} + \sum_{k \in K_*} \sum_{i=1}^{N_k} T_{(k,1),0}^{(i)}(\nu) \stackrel{d}{=} \sum_{i=1}^M T_{(k,1),0}^{(i)}(\nu),$$

for $M \stackrel{d}{=} \text{Geo}(p_*) + I_{\{k_1 \in K_*\}}$ and for any $k \in K_*$, where the latter stochastic equality follows from (11) and the fact that the random variables $T_{(k,1),0}$ are identically distributed for all $k \in K^*$. Taking $X_i(\nu) := T_{(k,1),0}^{(i)}(\nu)$ and applying Propositions III.4-III.6 then completes the proof. Also, noting that $\mathbb{E}M = I_{\{k_1 \in K_*\}} + p^*/(1-p^*)$ and $\mathbb{E}T_{(k,1),0} = \nu^{L_*-1}/L^*$ yields the statement of Theorem IV.1. ■

The k -th branch, $k \neq k_2$, is called *weakly dominant* if $k \in K_*$, i.e. if $L_k \geq L_j$ for all $j \neq k_2$. Based on Theorem IV.2, we may distinguish two scenarios, depending on whether branch k_1 is weakly dominant or not.

Suppose that branch k_1 is weakly dominant, i.e. $L_{k_1} = L_*$. In this case $M \stackrel{d}{=} \text{Geo}(p_*) + 1$ and so Theorem IV.2 implies that for every $0 < l_1 \leq L_{k_1}$ and $0 < l_2 \leq L_{k_2}$

$$\frac{T_{(k_1,l_1),(k_2,l_2)}(\nu)}{\mathbb{E}T_{(k_1,l_1),(k_2,l_2)}(\nu)} \xrightarrow{d} \text{Exp}(1), \quad \nu \rightarrow \infty.$$

Thus in this case the scaled transition time converges to an exponential random variable with unit mean as $\nu \rightarrow \infty$.

Suppose instead that branch k_1 is not weakly dominant, i.e. $L_{k_1} < L_*$. In this case $M \stackrel{d}{=} \text{Geo}(p_*)$ and so Theorem IV.2 implies that for every $0 < l_1 \leq L_{k_1}$ and $0 < l_2 \leq L_{k_2}$

$$\frac{T_{(k_1,l_1),(k_2,l_2)}(\nu)}{\mathbb{E}T_{(k_1,l_1),(k_2,l_2)}(\nu)} \xrightarrow{d} \frac{1-p_*}{p_*} \sum_{i=1}^{\text{Geo}(p_*)} Y_i, \quad \nu \rightarrow \infty,$$

where Y_i are independent and exponentially distributed random variables with unit mean.

V. MIXING TIMES

In the previous sections we have analyzed the transient behavior of our Markov process $(X_t)_{t \geq 0}$ in terms of hitting times. In this section we turn attention to the long-run behavior

of our Markov process and in particular examine the rate of convergence to the stationary distribution. We measure the rate of convergence in terms of the total variation distance and explore the intimate connection between the hitting times and the so-called *mixing time* of our Markov process. The mixing time describes the time required for the distance to stationarity to become small. It turns out that the mixing time is largely determined by the time it takes the process to escape from the second-longest branch in the state space, as formalized in Theorem V.1 below. In this section, we assume without loss of generality that the branches are indexed such that $L_1 \geq L_2 \geq \dots \geq L_K$, and denote by ∂_k the leaf node of the k -th branch. Also, we attach the starting state $X(0) = x \in \Omega$ as a superscript to our notation for the Markov process, and thus write $(X_t^x)_{t \geq 0}$.

A. Main result

Our objective is to bound the maximal distance over $x \in \Omega$, measured in terms of total variation, between the distribution at time t and the stationary distribution:

$$d(t) := \max_{x \in \Omega} \|\mathbb{P}\{X_t^x \in \cdot\} - \pi\|_{\text{TV}}.$$

We define the mixing time of our process as

$$t_{\text{mix}}(\epsilon, \nu) = \inf\{t \geq 0 : d(t) \leq \epsilon\}.$$

Theorem V.1. *The mixing time of the Markov process $(X_t)_{t \geq 0}$ satisfies*

$$t_{\text{mix}}(\epsilon, \nu) = \Theta(\nu^{L_2-1}),$$

i.e. $\exists C_1(\epsilon), C_2(\epsilon) > 0 \exists \nu_0 > 0$ such that for all $\nu > \nu_0$

$$C_1(\epsilon) \nu^{L_2-1} \leq t_{\text{mix}}(\epsilon, \nu) \leq C_2(\epsilon) \nu^{L_2-1}.$$

Theorem V.1 shows that it can take an extremely long time for the process $(X_t)_{t \geq 0}$ to reach stationarity, especially when ν is large. Such a long mixing time is typically due to the process being stuck for a considerably period in one of the components, and thus not visiting the states in the other components. This is particularly relevant when in the network has two or more dominant components which together attract the entire probability mass in the limit as $\nu \rightarrow \infty$. Indeed, in this case the mixing time provides an indication how long it takes for a certain fairness among the dominant components to occur. We will prove Theorem V.1 by deriving an upper bound for $d(t)$ using coupling in Subsection V-B and a matching lower bound for $d(t)$ using the bottleneck ratio and the notion of conductance in Subsection V-C.

B. Upper bound using coupling

It can be easily established that

$$d(t) \leq \bar{d}(t) := \max_{x, y \in \Omega} \|\mathbb{P}\{X_t^x \in \cdot\} - \mathbb{P}\{X_t^y \in \cdot\}\|_{\text{TV}}.$$

We consider $\bar{d}(t)$ instead of $d(t)$, because it can be bounded using a standard coupling technique. Consider all couplings of the processes (X_t^x, X_t^y) with the property that both (X_t^x) and (X_t^y) are Markov processes that have the same generator. Moreover, assume that the coupling is such that the two

processes stay together at all times after they have met for the first time. Under these assumptions, we have that

$$\begin{aligned} & \|\mathbb{P}\{X_t^x \in \cdot\} - \mathbb{P}\{X_t^y \in \cdot\}\|_{\text{TV}} \\ &= \max_{A \subseteq \Omega} |\mathbb{P}\{X_t^x \in A\} - \mathbb{P}\{X_t^y \in A\}| \\ &= \max_{A \subseteq \Omega} \mathbb{P}\{X_t^x \in A, X_t^y \notin A\} \\ &\leq \mathbb{P}\{X_t^x \neq X_t^y\} = \mathbb{P}\{\tau_{\text{couple}}^{x,y} > t\}, \end{aligned}$$

where the *coupling time* $\tau_{\text{couple}}^{x,y} = \min\{t \geq 0 : X_t^x = X_t^y\}$ denotes the first time the two processes (X_t^x) and (X_t^y) meet. Therefore, $d(t) \leq \bar{d}(t) \leq \max_{x, y \in \Omega} \mathbb{P}\{\tau_{\text{couple}}^{x,y} > t\}$, and the strength of this coupling inequality depends of course heavily on the choice of the coupling.

We now introduce a birth-and-death process $(M_t)_{t \geq 0}$ that will play a crucial role in our coupling. Let $(M_t)_{t \geq 0}$ describe the position of a particle that lives only on the two longest branches of the state space $\mathcal{C}_1 \cup \{0\} \cup \mathcal{C}_2$, starts in the leaf node ∂_2 , and moves within the two branches according to the same transition rates as $(X_t)_{t \geq 0}$. Call the particle that moves according to $(M_t)_{t \geq 0}$ particle 0. Consider also a particle 1 and particle 2 whose positions are governed by the coupled Markov process (X_t^x, X_t^y) . To be more specific, we denote the exact position of particle i at time t by (K_t^i, L_t^i) , $i = 0, 1, 2$, with K_t^i the branch and L_t^i the level.

Proposition V.2. *For all $x, y \in \Omega$, the coupling time $\tau_{\text{couple}}^{x,y}$, with $x = (k_1, l_1)$ and $y = (k_2, l_2)$ is stochastically bounded from above by the absorption time $T_{\partial_2, \partial_1}$ of state ∂_1 of the continuous-time birth-and-death process $(M_t)_{t \geq 0}$ starting from state $M_0 = \partial_2$.*

Proof: The coupling is such that particles 1 and 2 stay together at all times after they have met for the first time. Before that time, whenever a particle i ($i = 1, 2$) resides in $\mathcal{C}_1 \cup \mathcal{C}_2$, it is coupled to particle 0 in the following way: whenever particle 0 moves towards the root, particle i moves towards the root. In order to construct this coupling, we introduce a Poisson clock with rate L_1 . When the clock ticks, we first generate a $[0, 1]$ uniform random variable U and then do the following: move particle i down one level (towards the root) if $U < L_t^i/L_1$; also move particle 0 down one level if $U < L_t^0/L_1$. This coupling ensures that when $L_t^0 = L_t^i$ ($i = 1, 2$) and particle 0 moves down one level, then so does particle i . A consequence of the above coupling is that whenever particle 0 enters \mathcal{C}_1 both particle 1 and particle 2 reside in \mathcal{C}_1 . Extend the coupling by assuming that when particle 0 meets a particle i ($i = 1, 2$) in some state in \mathcal{C}_1 , the two particles keep making the same transitions as long as they are in \mathcal{C}_1 . Hence, inevitably, by the time particle 0 reaches ∂_1 , the particles 1 and 2 are coupled. ■

Since, for every $\nu > 0$ and for all $x, y \in \Omega$, $\tau_{\text{couple}}^{x,y}(\nu) \leq_{\text{st}} T_{\partial_2, \partial_1}(\nu)$, we arrive at the following result:

$$d(t) \leq \bar{d}(t) \leq \mathbb{P}\{T_{\partial_2, \partial_1}(\nu) > t\} \leq \frac{\mathbb{E}T_{\partial_2, \partial_1}(\nu)}{t}. \quad (12)$$

It then follows immediately that $t_{\text{mix}}(\epsilon, \nu) \leq \epsilon^{-1} \mathbb{E}T_{\partial_2, \partial_1}(\nu)$.

Using Theorem IV.1, we thus obtain the following upper bounds on the distance to stationarity and the mixing time.

Proposition V.3. *For the Markov process $(X_t)_{t \geq 0}$ the maximal total variation distance is bounded by*

$$d(t) \leq \frac{\mathbb{E}T_{\partial_2, \partial_1}(\nu)}{t} \sim \frac{1}{t} \left(\frac{L_1 + L_2}{L_1 L_2} \right) \nu^{L_2-1}, \quad \nu \rightarrow \infty,$$

and the mixing time is bounded by

$$t_{\text{mix}}(\epsilon, \nu) \leq \frac{1}{\epsilon} \mathbb{E}T_{\partial_2, \partial_1}(\nu) \sim \frac{1}{\epsilon} \left(\frac{L_1 + L_2}{L_1 L_2} \right) \nu^{L_2-1}, \quad \nu \rightarrow \infty.$$

C. Lower bound exploiting the bottleneck

Consider the activity process $(X_t)_{t \geq 0}$ with activation rate ν . For $S \subseteq \Omega$, let $\pi(S) := \sum_{(k,l) \in S} \pi_{(k,l)}(\nu)$ be the stationary probability of S . Define the *flow rate out of S* as $Q(S, S^c) := \sum_{(k,l) \in S, (j,m) \in S^c} \pi_{(k,l)}(\nu) q((k,l), (j,m))$ and the *conductance of S* as $\Phi(S) := Q(S, S^c)/\pi(S)$.

The conductance of the process $(X_t)_{t \geq 0}$ is defined as

$$\Phi_* := \min_{S: \pi(S) \leq 1/2} \Phi(S).$$

All the quantities we just defined depend on ν , but we suppress it for conciseness.

The following result, valid for any Markov process on a finite state space Ω with conductance Φ_* , shows how the conductance of the process yields a lower bound on the mixing time. It is a continuous-time version of Theorem 7.3 in [27] and since the proof is quite similar, it is omitted.

Lemma V.4. *For $\epsilon \in (0, \frac{1}{4})$, $t_{\text{mix}}(\epsilon) \geq (\frac{1}{2} - 2\epsilon) \frac{1}{\Phi_*}$.*

We now exploit the fact that our activity process on the state space Ω has a geometric feature usually referred to as *bottleneck*, that strongly influences the mixing time. This bottleneck indeed makes some parts of Ω difficult to reach, resulting in a small conductance. As it turns out, \mathcal{C}_2 will be the bottleneck.

Proposition V.5. *The conductance of \mathcal{C}_2 satisfies*

$$\Phi(\mathcal{C}_2) \sim L_2 \nu^{1-L_2}, \quad \nu \rightarrow \infty,$$

and hence, for $\epsilon \in (0, \frac{1}{4})$,

$$t_{\text{mix}}(\epsilon, \nu) \geq \left(\frac{1}{2} - 2\epsilon \right) \frac{1}{L_2} \nu^{L_2-1}, \quad \nu \rightarrow \infty. \quad (13)$$

Proof: Since $\mathcal{C}_1 \cup \{0\}$ will have at least half of the probability mass for ν sufficiently large, it is clear that $\pi(\mathcal{C}_2) \leq 1/2$ when $\nu \rightarrow \infty$. From (2) it follows that if l and m belong to the same component k of size L , then $\frac{\pi_{(k,m)}(\nu)}{\pi_{(k,l)}(\nu)} = \frac{l!(L-l)!}{m!(L-m)!} \nu^{m-l}$, as $\nu \rightarrow \infty$. Thus the conductance of \mathcal{C}_2 satisfies

$$\Phi(\mathcal{C}_2) = \frac{\pi_{(2,1)}(\nu) \cdot 1}{\sum_{l=1}^{L_2} \pi_{(2,l)}(\nu)} = \frac{\pi_{(2,1)}(\nu)}{\sum_{l=1}^{L_2} \pi_{(2,l)}(\nu)} \sim L_2 \nu^{1-L_2}.$$

Then Lemma V.4 gives the lower bound (13), since by definition $\Phi_* \leq \Phi(\mathcal{C}_2)$. ■

VI. CONCLUSION AND EXTENSIONS

We have examined transient throughput characteristics and associated starvation effects in CSMA networks in terms of the transition times between dominant activity states. We established how the magnitude of the transition time scales with the activation rate and the sizes of the various network components in partite interference graphs. We also proved that in several cases the scaled transition time has an asymptotically exponential distribution and discussed the connection with related exponentiality results for rare events and meta-stability effects in statistical physics. In addition, we investigated the convergence rate to equilibrium of the activity process in terms of mixing times.

In the present paper we have focused on partite interference graphs with uniform activation rates, giving rise to a star-shaped state space, but most of the methods and results extend to more general networks, as long as there is a unique path between any two activity states. For example, we could allow for the branches to be general trees, with $M_{k,l}$ nodes in the k -th tree at distance l from the root, and transition rates $b_{k,l} f_{k,l}(\nu)$ and $d_{k,l}$ away from and towards the root, respectively. The node-dependent functions $f_{k,l}(\nu)$ are particularly relevant, since taking $f_{k,l}(\nu) = \nu^{1/L_k}$ could for example serve to balance the long-term throughputs in the various components. In those cases, the expectation of the transition time will scale differently, but most of the distributional results for the scaled transition time continue to hold, even though qualitatively different scenarios arise, with some branches being visited for relatively short periods but an overwhelmingly large number of times.

REFERENCES

- [1] F. P. Kelly, *Reversibility and Stochastic Networks*. Wiley, 1979.
- [2] R. Boorstyn, A. Kershenbaum, B. Maglaris, and V. Sahin, "Throughput analysis in multihop CSMA packet radio networks," *IEEE Trans. Commun.*, vol. 35, no. 3, pp. 267–274, 1987.
- [3] F. P. Kelly, "Stochastic models of computer communication systems," *J. Roy. Statist. Soc. Ser. B*, vol. 47 (3), pp. 379–395, 1985.
- [4] A. Kershenbaum, R. Boorstyn, and M. Chen, "An algorithm for evaluation of throughput in multihop packet radio networks with complex topologies," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 6, pp. 1003–1012, 1987.
- [5] X. Wang and K. Kar, "Throughput modelling and fairness issues in CSMA/CA based ad-hoc networks," in *Proc. IEEE Infocom*, 2005, pp. 23–34.
- [6] M. Durvy and P. Thiran, "A packing approach to compare slotted and non-slotted medium access control," in *Proc. IEEE Infocom*, 2006, pp. 1–12.
- [7] J. Shi, O. Gurewitz, V. Mancuso, J. Camp, and E. W. Knightly, "Measurement and modeling of the origins of starvation in congestion controlled mesh networks," in *Proc. IEEE Infocom*, 2008, pp. 1633–1641.
- [8] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, 2000.
- [9] S. Liew, C. Kai, H. Leung, and P. Wong, "Back-of-the-envelope computation of throughput distributions in CSMA wireless networks," *IEEE Trans. Mob. Comput.*, vol. 9, no. 9, pp. 1319–1331, 2010.
- [10] Y. Suhov and U. A. Rozikov, "A hard-core model on a Cayley tree: an example of a loss network," *Queueing Systems*, vol. 46, no. 1, pp. 197–212, 2004.
- [11] S. Zachary and I. Ziedins, "Loss networks and Markov random fields," *J. Appl. Prob.*, vol. 36, no. 2, pp. 403–414, 1999.

- [12] R. Gouet and A. Sudbury, "Blocking and dimer processes on the Cayley tree," *J. Stat. Phys.*, vol. 130, no. 5, pp. 935–955, 2008.
- [13] O. Häggström, "Ergodicity of the hard-core model on \mathbb{Z}^2 with parity-dependent activities," *Ark. Mat.*, vol. 35, no. 1, pp. 171–184, 1997.
- [14] P. M. van de Ven, J. S. H. van Leeuwen, D. Denteneer, and A. J. E. M. Janssen, "Spatial fairness in linear random-access networks," *Performance Evaluation*, vol. 69, no. 3–4, pp. 121–134, 2012.
- [15] F. den Hollander, "Metastability under stochastic dynamics," *Stoch. Proc. Appl.*, vol. 114, no. 1, pp. 1–26, 2004.
- [16] E. Olivieri and E. Scoppola, "Markov chains with exponentially small transition probabilities: first exit problem from a general domain. I. The reversible case," *J. Stat. Phys.*, vol. 79, no. 3–4, pp. 613–647, 1995.
- [17] S. Karlin and J. McGregor, "Coincidence properties of birth and death processes," *Pacific J. Math.*, vol. 9, pp. 1109–1140, 1959.
- [18] J. Keilson, "A review of transient behavior in regular diffusion and birth-death processes. Part II," *J. Appl. Prob.*, vol. 2, no. 2, pp. 405–428, 1965.
- [19] —, "Log-concavity and log-convexity in passage time densities of diffusion and birth-death processes," *J. Appl. Prob.*, vol. 8, no. 2, pp. 391–398, 1971.
- [20] D. J. Aldous, "Markov chains with almost exponential hitting times," *Stoch. Proc. Appl.*, vol. 13, no. 3, pp. 305–310, 1982.
- [21] D. J. Aldous and M. Brown, "Inequalities for rare events in time-reversible Markov chains I," in *Stochastic Inequalities*, ser. IMS Lecture Notes Monogr. Ser., 1992, vol. 22, pp. 1–16.
- [22] —, "Inequalities for rare events in time-reversible Markov chains II," *Stoch. Proc. Appl.*, vol. 44, no. 1, pp. 15–25, 1993.
- [23] R. Grübel and M. Reich, "Rarity and exponentiality: an extension of Keilson's theorem, with applications," *J. Appl. Prob.*, vol. 42, no. 2, pp. 393–406, 2005.
- [24] J. Keilson, "A limit theorem for passage times in ergodic regenerative processes," *Ann. Math. Statist.*, vol. 37, no. 4, pp. 866–870, 1966.
- [25] —, *Markov Chain Models—Rarity and Exponentiality*. Springer, 1979.
- [26] W. Ledermann and G. Reuter, "Spectral theory for the differential equations of simple birth and death processes," *Philos. Trans. Roy. Soc. London. Ser. A*, vol. 246, pp. 321–369, 1954.
- [27] D. Levin, Y. Peres, and E. Wilmer, *Markov Chains and Mixing Times*. AMS, 2009.

APPENDIX

A. Proof of Lemma II.3

Order the state space as $\Omega = \{L, L-1, \dots, 1, 0\}$ and consider the generator matrix $Q(\nu)$ of the process $(X_t)_{t \geq 0}$ with 0 an absorbing state. That is,

$$Q(\nu) = \begin{bmatrix} q_L(\nu) & L & 0 & & & \\ \nu & q_{L-1}(\nu) & L-1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & (L-1)\nu & q_1(\nu) & 1 & \\ & & 0 & 0 & 0 & \end{bmatrix},$$

where the diagonal elements are $q_l(\nu) = -(L-l)\nu - l$ for $l = 1, \dots, L$. Write $Q(\nu)$ as

$$Q(\nu) = \begin{pmatrix} \mathbf{T}(\nu) & \mathbf{t}(\nu) \\ \mathbf{0} & 0 \end{pmatrix},$$

where $\mathbf{T}(\nu)$ is an $L \times L$ invertible matrix. Since the characteristic polynomials of $-Q(\nu)$ and $-\mathbf{T}(\nu)$ satisfy the relation $p_{-Q(\nu)}(z) = -z p_{-\mathbf{T}(\nu)}(z)$, the spectrum of $-Q(\nu)$ consists of that of $-\mathbf{T}(\nu)$ plus the eigenvalue zero with multiplicity one. Denote by $D(\nu)$ the $L \times L$ diagonal matrix, whose diagonal entries are $\{\sqrt{\theta_l(\nu)}\}_{l=L}^1$, where the θ 's are the so-called *potential coefficients*, defined as $\theta_L(\nu) = 1$ and $\theta_{l-1}(\nu) = \frac{l}{(L-l)\nu} \theta_l(\nu)$. The $L \times L$ matrix $G(\nu) = -D(\nu)^{1/2} \mathbf{T}(\nu) D(\nu)^{-1/2}$ is tridiagonal and symmetric with diagonal entries $g_{l,l}(\nu) = q_{L-l+1}(\nu)$ and $g_{l,l+1}(\nu) =$

$g_{l+1,l}(\nu) = -\sqrt{l \cdot (L-l+1)\nu}$. Since $G(\nu)$ is similar to $-\mathbf{T}(\nu)$, they have the same spectrum. Denote by $\mathcal{D}(p, R)$ the closed disc centered in p with radius R , i.e. $\mathcal{D}(p, R) = \{z \in \mathbb{C} : |z - p| \leq R\}$. Consider the *Gershgorin discs* $\{\mathcal{D}_l(\nu)\}_{l=1}^L$ of $G(\nu)$, defined as $\mathcal{D}_l(\nu) := \mathcal{D}(-q_l(\nu), R_l(\nu))$, where the radius $R_l(\nu)$ is the sum of the absolute values of the non-diagonal entries in the $L-l+1$ -th row, i.e. $R_l(\nu) := \sum_{m \neq L-l+1} |g_{L-l+1,m}(\nu)|$. Then

$$\mathcal{D}_L(\nu) = \mathcal{D}(L, \sqrt{L\nu}),$$

$$\mathcal{D}_{L-1}(\nu) = \mathcal{D}(L-1+\nu, \sqrt{L\nu} + \sqrt{2(L-1)\nu}),$$

...

$$\mathcal{D}_2(\nu) = \mathcal{D}(2 + (L-2)\nu, \sqrt{3(L-2)\nu} + \sqrt{2(L-1)\nu}),$$

$$\mathcal{D}_1(\nu) = \mathcal{D}(1 + (L-1)\nu, \sqrt{2(L-1)\nu}).$$

We now exploit the second Gershgorin circle theorem, which is reproduced here for completeness.

Theorem. *If the union of j Gershgorin discs of a real $r \times r$ matrix A is disjoint from the union of the other $r-j$ Gershgorin discs, then the former union contains exactly j and the latter the remaining $r-j$ eigenvalues of A .*

In our case, for ν sufficiently large, the disc $\mathcal{D}_L(\nu)$ does not intersect with the union $\bigcup_{l=1}^{L-1} \mathcal{D}_l(\nu)$, thus the smallest eigenvalue $\alpha_1(\nu)$ lies in $\mathcal{D}_L(\nu)$ and the other $L-1$ ones in $\bigcup_{l=1}^{L-1} \mathcal{D}_l(\nu)$. Hence, for ν sufficiently large, $\alpha_1(\nu) \leq L + \sqrt{L\nu}$ and $\alpha_i(\nu) \geq (L-1) + \nu - \sqrt{\nu}(\sqrt{L} + \sqrt{2(L-1)})$ for $i = 2, \dots, L$. Therefore, $0 < \frac{\alpha_1(\nu)}{\alpha_i(\nu)} \leq \frac{L + \sqrt{L\nu}}{\nu - \sqrt{\nu}(\sqrt{L} + \sqrt{2(L-1)})}$, and so $\lim_{\nu \rightarrow \infty} \alpha_1(\nu)/\alpha_i(\nu) = 0$ for $i = 2, \dots, L$. Hence,

$$\mathbb{E}T_{L,0}(\nu) \cdot \alpha_1(\nu) = 1 + \sum_{i=2}^L \frac{\alpha_1(\nu)}{\alpha_i(\nu)} \rightarrow 1, \quad \nu \rightarrow \infty,$$

while for $2 \leq i \leq L$,

$$\mathbb{E}T_{L,0}(\nu) \cdot \alpha_i(\nu) > \frac{\alpha_i(\nu)}{\alpha_1(\nu)} \rightarrow \infty, \quad \nu \rightarrow \infty.$$

B. Proof of Proposition III.6

Using $\mathbb{E}S_M(\nu) = \mathbb{E}M \cdot \mathbb{E}X(\nu)$, gives

$$\frac{S_M(\nu)}{\mathbb{E}S_M(\nu)} = \frac{\sum_{i=1}^M X_i(\nu)}{\mathbb{E}M \mathbb{E}X(\nu)} = \frac{1}{\mathbb{E}M} \sum_{i=1}^M \frac{X_i(\nu)}{\mathbb{E}X_i(\nu)}.$$

Thus the Laplace transform of $S_M/\mathbb{E}S_M$ may be written as

$$\begin{aligned} \mathcal{L}_{S_M(\nu)/\mathbb{E}S_M(\nu)}(s) &= \mathcal{L}_{\sum_{i=1}^M X_i(\nu)/\mathbb{E}X_i(\nu)}\left(\frac{s}{\mathbb{E}M}\right) \\ &= G_M\left(\mathcal{L}_{X_i(\nu)/\mathbb{E}X_i(\nu)}\left(\frac{s}{\mathbb{E}M}\right)\right), \end{aligned}$$

where $G_M(z) = \mathbb{E}(z^M)$. The assumption $X(\nu)/\mathbb{E}X(\nu) \xrightarrow{d} Y$ implies that $\lim_{\nu \rightarrow \infty} \mathcal{L}_{X_i(\nu)/\mathbb{E}X_i(\nu)}(t) = \mathcal{L}_Y(t)$ for all $t \geq 0$ and $i \in \mathbb{N}$. Hence

$$\begin{aligned} \lim_{\nu \rightarrow \infty} \mathcal{L}_{S_M(\nu)/\mathbb{E}S_M(\nu)}(s) &= \lim_{\nu \rightarrow \infty} G_M\left(\mathcal{L}_{X_i(\nu)/\mathbb{E}X_i(\nu)}\left(\frac{s}{\mathbb{E}M}\right)\right) \\ &= G_M\left(\mathcal{L}_Y\left(\frac{s}{\mathbb{E}M}\right)\right), \end{aligned}$$

which is the Laplace transform of $\frac{1}{\mathbb{E}M} \sum_{i=1}^M Y_i$. Invoking the continuity theorem for Laplace transforms concludes the proof.